Insights into salt tolerance from the genome of *Thellungiella salsuginea*

Hua-Jun Wu[a,1], Zhonghui Zhang[a,1], Jun-Yi Wang[b,1], Dong-Ha Oh[c,1], Maheshi Dassanayake[c,1], Binghang Liu[b,1], Quanfei Huang[b,1], Hai-Xi Sun[a], Ran Xia[a], Yaorong Wu[a], Yi-Nan Wang[a], Zhao Yang[a], Yang Liu[a], Wanke Zhang[a], Huawei Zhang[a], Jinfang Chu[a], Cunyu Yan[a], Shuang Fang[a], Jinsong Zhang[a], Yiqin Wang[a], Fengxia Zhang[a], Guodong Wang[a], Sang Yeol Lee[d], John M. Cheeseman[c], Bicheng Yang[b], Bo Li[b], Jiumeng Min[b], Linfeng Yang[b], Jun Wang[b,2], Chengcai Chu[a,2], Shou-Yi Chen[a,2], Hans J. Bohnert[c,d,e], Jian-Kang Zhu[f,g,2], Xiu-Jie Wang[a,2], and Qi Xie[a,2]

[a]State Key Laboratory of Plant Genomics, National Center for Plant Gene Research (Beijing), Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China; [b]BGI-Shenzhen, Shenzhen 518083, China; [c]Department of Plant Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801; [d]Division of Applied Life Sciences, Gyeongsang National University, Jinju 660-701, Korea; [e]College of Science, King Abdulaziz University, Jeddah 21589, Saudi Arabia; [f]Shanghai Center for Plant Stress Biology and Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China; and [g]Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, IN 47907

Contributed by Jian-Kang Zhu, June 15, 2012 (sent for review February 27, 2012)

***Thellungiella salsuginea*, a close relative of *Arabidopsis*, represents an extremophile model for abiotic stress tolerance studies. We present the draft sequence of the *T. salsuginea* genome, assembled based on ~134-fold coverage to seven chromosomes with a coding capacity of at least 28,457 genes. This genome provides resources and evidence about the nature of defense mechanisms constituting the genetic basis underlying plant abiotic stress tolerance. Comparative genomics and experimental analyses identified genes related to cation transport, abscisic acid signaling, and wax production prominent in *T. salsuginea* as possible contributors to its success in stressful environments.**

genome sequence | halophyte | gene duplication | stress response

Abiotic stresses such as salinity, drought, or temperature extremes greatly impair plant growth and development and crop yield. The need to cultivate marginal lands to increase food production in the future will expose crops to adverse conditions and exacerbate agricultural problems. Thus, enormous value will come from a better understanding of the mechanisms through which plant tolerance of abiotic stresses is achieved. Most studies on plant response mechanisms leading to stress tolerance have been conducted with the model plant *Arabidopsis*, which has a relatively low capacity to survive abiotic stresses. However, the *Arabidopsis* model and work on a variety of other species have provided clues about enhanced stress tolerance based on individual genes in a number of pathways. Unfortunately, in nearly all cases, genes with a stress-alleviating quality under controlled conditions have failed to generate stress protection in the field. This lack of success argues for developing models that can provide crucial insights into mechanisms that confer high levels of stress tolerance in species that exhibit natural tolerance (1, 2).

The crucifer *Thellungiella salsuginea* (Pallas), a close relative of *Arabidopsis* originally classified as *Thellungiella halophila*, is a halophyte with exceptionally high resistance to cold, drought, and oxidative stresses as well as salinity (1–6). *T. salsuginea* is exemplary by its short life cycle, self-fertility, and being genetically transformable (3). These characteristics make the species an excellent model for unraveling the factors that constitute abiotic stress tolerance (1–8). Further advantages are its relatively small genome size [approximately twice that of *A. thaliana* (3)] and the availability of ecotypes that show a range of stress responses (8).

High-throughput studies of *T. salsuginea* thus far have been restricted largely to the characterization of its transcriptome (5, 7, 8). In addition, comparisons of transcriptome stress responses in *T. salsuginea* and *Arabidopsis thaliana* highlighted different regulation of well-known pathways as well as unstudied stress-related genes (4, 6). The recent publication of the genome sequence of the congeneric species *Thellungiella parvula* has enabled consideration of the genomic and evolutionary basis of stress adaptation with the improved resolution provided by a comparative approach (9).

Here we present the genome sequence and overall chromosome structure of *T. salsuginea* and use comparative genomics and experimental approaches to identify genes in *T. salsuginea* that contribute to its success in stressful environments.

Results

Sequence and Assembly. We sequenced the genome of *T. salsuginea* (Shandong ecotype) using the paired-end Solexa sequencing method (Illumina GA II system). Based on flow cytometry of isolated nuclei stained with propidium iodide (3), we expected a genome size of ~260 Mb (*SI Appendix*, Table S1). Thus, with a total of 34.8 Gb of high-quality sequences, the genome was covered ~134-fold (*SI Appendix*, Table S2). The final length of the assembled sequences amounted to ~233.7 Mb, covering about 90% of the estimated genome size. The assembly consists of 2,682 scaffolds, the 10 longest of which range from 1.9–6.8 Mb (*SI Appendix*, Table S1) and represent 17% of the assembled genome.

In the absence of genetic and physical markers, we assigned many remaining scaffolds to blocks (chromosome segments) identified by Lysak and coworkers (10, 11) by comparative chromosome painting, which represents the ancestral karyotype in the crucifers. By tracing these blocks, we anchored 515 scaffolds onto seven chromosomes, with a total size of 186 Mb (about 80% of the total assembled genome; Fig. 1).

Repetitive Sequences. The size of the *T. salsuginea* genome is approximately twice that of *A. thaliana*, largely reflecting a proliferation of transposable elements (TEs). A repetitive-sequence

Author contributions: J.W., C.C., S.-Y.C., J.-K.Z., X.-J.W., and Q.X. designed research; Z.Z., J.-Y.W., and B.Y. performed research; H.-J.W., Z.Z., D.-H.O., M.D., B. Liu, Q.H., H.-X.S., R.X., Y.Wu, Y.-N.W., Z.Y., Y.L., W.Z., H.Z., J.C., C.Y., S.F., J.Z., Y.W., F.Z., G.W., B. Li, J.M., and L.Y. analyzed data; and H.-J.W., Z.Z., S.Y.L., J.M.C., H.J.B., X.-J.W., and Q.X. wrote the paper.

The authors declare no conflict of interest.

Data deposition: The sequence for the *Thellungiella salsuginea* genome reported in this paper has been deposited in the Data Bank of Japan (DDBJ)/European Molecular Biology Laboratory (EMBL)/GenBank database, http://www.ncbi.nlm.nih.gov/bioproject/?term=txid72664 (accession no. AHIU00000000; PID 80723).

[1]H.-J.W., Z.Z., J.-Y.W., D.-H.O., M.D., B. Liu, and Q.H. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: wangj@genomics.org.cn, ccchu@genetics.ac.cn, sychen@genetics.ac.cn, jkzhu@purdue.edu, xjwang@genetics.ac.cn, or qxie@genetics.ac.cn.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1209954109/-/DCSupplemental.

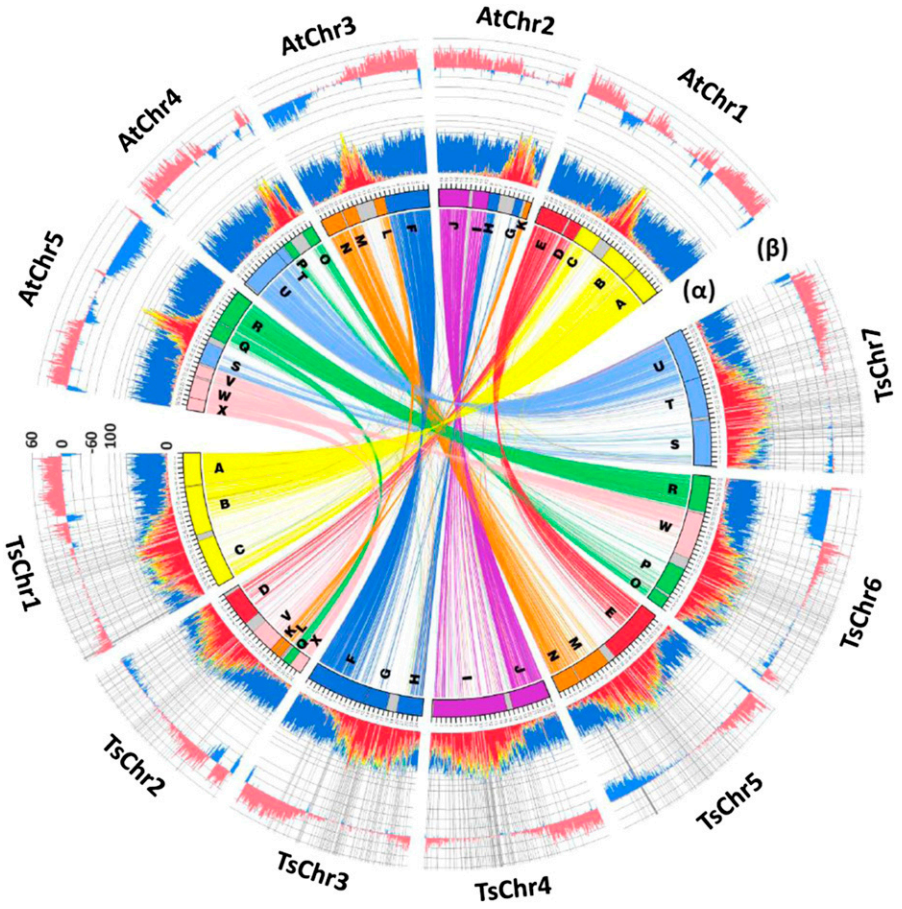

Fig. 1. The genome of *T. salsuginea*. The assembled seven chromosomes of *T. salsuginea* are shown in a comparison with *A. thaliana*. Ancestral karyotype blocks A–X (10) are shown in different colors. Sequences with >70% similarity over the length of 2 kb are connected by links of the same colors as the ancestral karyotype blocks. Histogram α represents the distribution of TEs and predicted genes. Class I retrotransposons, class II DNA transposons, and unclassified repetitive sequences are indicated by red, orange, and yellow colors, respectively, and the predicted genes are shown in blue. The outer histogram β shows the percentage of sequences that can be aligned between the two species with >70% identity. Alignments longer than 500 bp were counted, and their percentages per 100-Kb windows are presented, with the alignments in opposite directions in the two genomes shown in blue and the alignments in the same direction shown in pink. Scales in the *y*-axes of the histograms are in percentage. Radial lines indicate the boundaries of the scaffolds used in the *T. salsuginea* genome assembly.

database search combined with detection of TEs identified 121 Mb of repetitive sequences (*SI Appendix*) representing ~52% of the genome (*SI Appendix*, Tables S1 and S3). This percentage is much higher than the 13.2% and 7.5% TE contents of *A. thaliana* (12) and *T. parvula* (13), respectively. Like most of higher plant genomes, class I TEs (retrotransposons), especially LTR retrotransposons, account for a comparatively high percentage (36%) of the *T. salsuginea* genome. Among these, gypsy and copia are the two most abundant TE families.

Gene Space. A total of 28,457 protein-coding regions were predicted in the sequenced *T. salsuginea* genome using a combination of homologous sequence searches, *ab initio* gene predictions, and transcriptome data comparisons with the genome sequence (*SI Appendix*, Table S1 and Dataset S1). In addition, 447 tRNAs, 11 rRNAs, 432 snRNAs, and 162 microRNAs (including 126 conserved ones) were identified (*SI Appendix*, Tables S1 and S4). The overall ORF length distribution of *T. salsuginea* is comparable to that of *A. thaliana*, with a slightly higher proportion of ORFs shorter than 1,000 bp identified in *T. salsuginea* (*SI Appendix*, Fig. S1). The average exon length of *T. salsuginea* and *A. thaliana* genes is similar (228 and 224 bp, respectively), whereas the average intron length of *T. salsuginea* is ~30% larger than that of *A. thaliana* (200 and 157 bp, respectively) (*SI Appendix*, Table S1) (12).

About 93% of the predicted coding regions showed at least partial similarity with known protein sequences and can be annotated (*SI Appendix*, Table S1). Comparative genomic analysis identified 984 *T. salsuginea* unique gene families and 9,667 families shared by *T. salsuginea*, *A. thaliana*, *Carica papaya*, and *Vitis vinifera* (Fig. 2*A*). Consistent with their close evolutionary relationships, 16,358 gene families were shared by *T. salsuginea* and *A. thaliana*, representing 93.7% and 95.2% of all gene families, respectively (Fig. 2*A*). The protein-coding gene models were compared with *A. thaliana* and *T. parvula* (13), and orthologous gene models were identified. *Thellungiella* species share comparable numbers of orthologs with each other and with *A. thaliana*. Both *Thellungiella* species contain large numbers of "orphan" genes for which no orthologs exist in *A. thaliana* (Fig. 2*B*, indicated by red color). Among all orphan gene models, 54.7%, 62.8%, and 36.5% in *T. salsuginea*, *T. parvula*, and *A. thaliana*, respectively, lacked any Gene Ontology (GO) annotation and hence are annotated as functionally unknown (Dataset S2).

Evolutionary History. Phylogenetic analyses (*SI Appendix*, Fig. S2) indicate a time of divergence between *T. salsuginea* and *A. thaliana* of 7–12 Mya, following the split of the *Arabidopsis* and *Brassica* lineages (14). A similar time has been suggested for the *A. thaliana* and *T. parvula* split (9). Previous studies have suggested that the *A. thaliana* genome shows signatures of the paleohexaploidy whole-genome duplication (WGD) event γ proposed at the base of eudicot divergence and two recent WGD events, β and α, within the crucifer lineage (14). Similarly, two peaks representing the β and α events [~0.28 and ~0.6 fourfold degenerative third-codon transversion (4dTv) distance] were identified in *T. salsuginea* (*SI Appendix*, Fig. S3), suggesting that the divergence of *T. salsuginea* and *A. thaliana* occurred after the two most recent WGD events.

Tandem duplication, segmental duplication, and retrotransposition-directed duplications (*SI Appendix*) were analyzed to weigh their contribution to the variation in gene copy number and to probe for a possible bias in gene functional enrichment in *T. salsuginea* and *A. thaliana*. The total numbers of tandemly

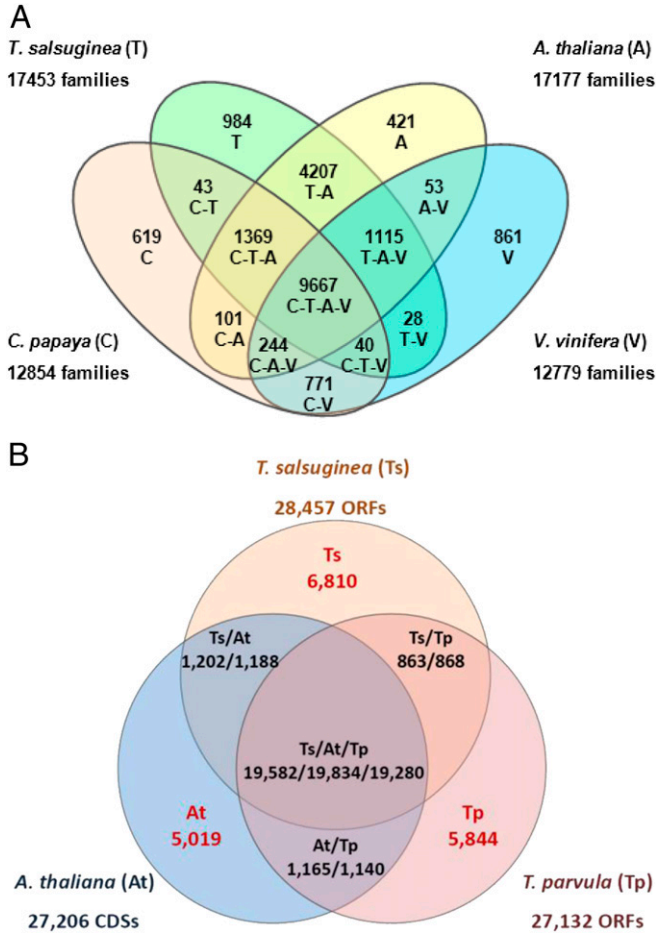

Fig. 2. Comparison of orthologous genes and gene groups. (*A*) Shared orthologous gene clusters among the *T. salsuginea*, *A. thaliana*, *C. papaya*, and *V. vinifera* genomes. Program OrthoMCL was applied to identify orthologous groups among the *T. salsuginea* (T), *A. thaliana* (A), *C. papaya* (C), and *V. vinifera* (V) genomes. (*B*) Shared orthologous genes among crucifers *T. salsuginea* (Ts), *T. parvula* (Tp), and *A. thaliana* (At). Orthologs were identified using OrthoMCL. Genes from different species were considered as orthologs if the shared homology in their deduced amino acid sequences (BlastP, e < 0.00001) was more than 50% of the size of the genes being compared. Numbers of orphan genes lacking an ortholog in the other two species are shown in red. Lists of genes and their GO annotations in each category are given in Dataset S2.

duplicated genes and segmentally duplicated genes are similar in the two species. In contrast, genes carrying an LTR transposon and retrogenes, although they only accounted for a small proportion of genes in total, were significantly more expanded in *T. salsuginea* (Fig. 3 *A* and *B*), probably reflecting the high abundance of TEs, especially LTR retrotransposons. About 30% of the tandemly duplicated genes and 60% of segmentally duplicated genes were shared between *T. salsuginea* and *A. thaliana*. Very few duplication events seemingly mediated by retrogene action, especially those based on LTRs, were conserved, indicating that those events occurred after the divergence of the two species. GO analysis revealed that both tandem and segmental duplications tended to affect genes in similar functional categories (*SI Appendix*, Table S5).

Genome Designed for Stress Response Capacity. GO terms were assigned to the *T. salsuginea* predicted ORFs and *A. thaliana* annotated genes using the Blast2GO pipeline (15). The term "response to stimulus" was identified among the GO categories that differed significantly between *T. salsuginea* and *A. thaliana*, with more genes in the *T. salsuginea* genome classified into this category (Fig. 4). This divergence represents the contribution

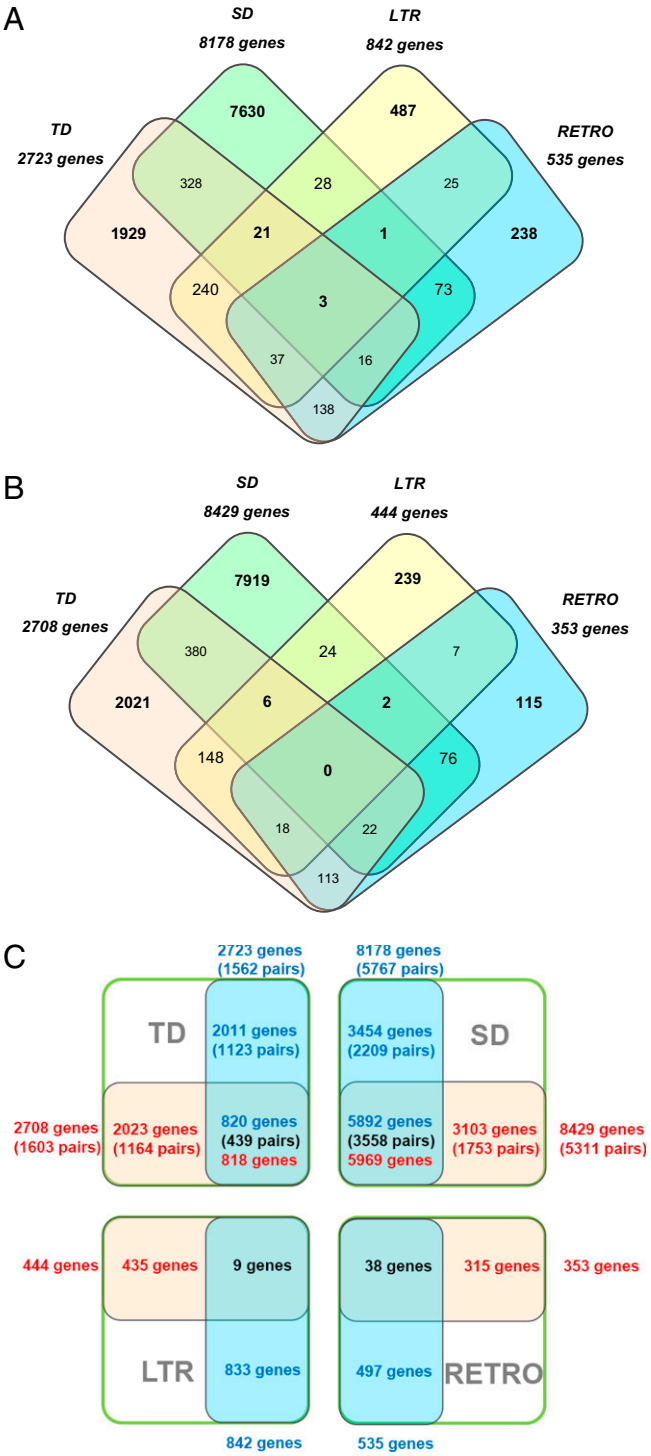

Fig. 3. Comparison of tandem duplication (TD), segmental duplication (SD), and retrotransposition (LTR and RETRO) events in the *T. salsuginea* and *A. thaliana* genomes. (*A* and *B*) Assembled venn plots show shared and specific genes among different types of gene duplications in *T. salsuginea* (*A*) and *A. thaliana* (*B*). (*C*) Comparisons of each type of gene duplication in the two species. Numbers of duplicated genes in *T. salsuginea* are in red; numbers in *A. thaliana* are in blue. LTR, LTR retrotransposon carrying genes; RETRO, retrogenes; SD, segmental duplicated genes; TD, tandem duplicated genes.

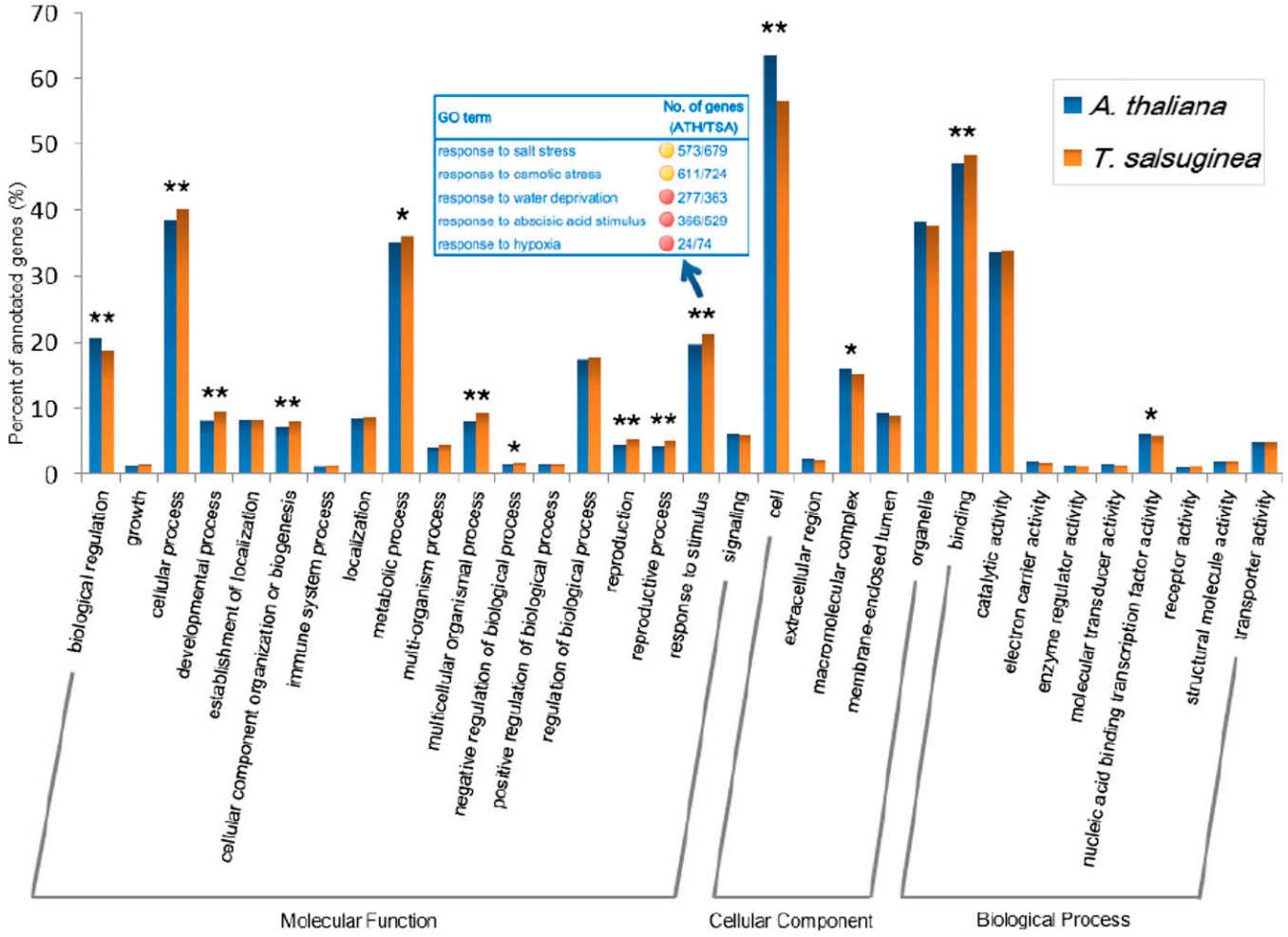

Fig. 4. GO comparison of *T. salsuginea* and *A. thaliana*. Blast2GO results of protein-coding regions from *T. salsuginea* and *A. thaliana* were mapped to categories in the second level of GO terms. Fisher's exact test was used to evaluate the significance of differences in GO category enrichment in the two species. GO terms that contain more than 1% of total genes were included in the graph; those with *P* values below 0.01 and 0.05 are marked by double stars and stars, respectively, on the histogram. Subcategories of the term "response to stimulus" that differ significantly in the two species are shown in the box.

of the combinatory effect of four major duplication events (*SI Appendix*, Table S5). The same trend also was observed in the *T. parvula* genome (13). Detailed analysis revealed that genes related to "response to salt stress," "osmotic stress," "water deprivation," "ABA stimulus," and "hypoxia" were expanded in the "response to stimulus" category in *T. salsuginea* compared with *A. thaliana* (Fig. 4). As a genome signature, this difference may be caused by and could contribute to the high salinity- and drought-tolerant phenotype of *T. salsuginea*.

A total of 21 transcription factor families were found to be expanded in the *T. salsuginea* genome compared with *A. thaliana* (*SI Appendix*, Table S6). These expansions may be associated with the adaptation of *T. salsuginea* to extreme environments, because individual members of some families in *A. thaliana* have been linked previously with stress resistance. For example, the *RAV* gene family, which had been reported to respond to high salt and cold stresses (16, 17), expanded from six members in the *A. thaliana* genome to nine in *T. salsuginea*. Other gene families with known functions in abiotic stress response that expanded in numbers in *T. salsuginea* include the *NF-X1*, *GRAS*, *HSF*, and *Trihelix* families. It has been shown that one *NF-X1* family member, *AtNFXL1*, is required for growth of *Arabidopsis* under salinity stress (18), and *RGL3* in the *GRAS* family can be up-regulated transiently by cold stress (19). *HSFA2*, the most abundant member of the heat-shock response *HSF* family, also is induced by salinity in *Arabidopsis*, and its overexpression enhances salt and osmotic stress tolerance (20). The *GTgamma* subfamily in the *Trihelix* family contains three genes induced by most abiotic stresses in rice (21). Overexpression of two soybean *Trihelix* family genes in *Arabidopsis* greatly enhanced salt, drought, and cold tolerance (22).

Expansion of Genes Related to the Maintenance of Ion Equilibrium. Effective establishment of ionic and osmotic equilibrium is important for plant salinity and drought tolerance. Comparison of gene families involved in ion transport in *T. salsuginea* and *A. thaliana* indicated that gene families providing ionic stress protection, including *HKT, CNGC*, *PPa*, *ACA*, *AVP*, *ATBGL*, *CIPK*, and *CDPK* (23–25), have more members in *T. salsuginea* (*SI Appendix*, Table S7). One group, the *HKT* gene family, encodes Na^+/K^+ transporters that may provide key components affecting or determining salt tolerance in plants (26–29). Recently, two *HKT1* transcripts have been reported in *T. salsuginea* (30); however, the genome annotation revealed a third homolog (*Ts6g08740/TsHKT1;3*). The three *TsHKT1* genes exist in a tandem gene array, similar to the tandem duplication of two *HKT1* genes in *T. parvula* (13); only one copy is present in *A. thaliana* (*SI Appendix*, Table S7). Based on phylogenetic analysis, *Ts6g08650/TsHKT1;1* is clustered with *AtHKT1,* whereas the

other two *TsHKT1* cluster with the two *TpHKT1* genes in another group (*SI Appendix*, Fig. S4*A*). All three *T. salsuginea HKT1* genes were found to be expressed, with the expression of *TsHKT1;2* (*Ts6g08730*) being significantly higher than the expression of the other two genes (*SI Appendix*, Fig. S4*B*).

Stress Tolerance-Supportive Genes and Pathways. Reduction of water loss by epicuticular wax is a strategy used by plants to defend themselves against abiotic stresses (31). Throughout its development, *T. salsuginea* exhibits highly glaucous leaves indicative of complex epicuticular wax organization. We found a tandem duplicated gene in *T. salsuginea* encoding cytochrome P450-dependent midchain hydroxylase MAH1/CYP96A15, which currently is the only known enzyme in the wax-producing–related alkane-forming pathway. This gene is also tandemly duplicated in the *T. parvula* genome (13) but is not duplicated in *A. thaliana* (*SI Appendix*, Fig. S5 and Table S8), perhaps explaining the previous finding that the wax content was much higher in *T. salsuginea* than in *A. thaliana* leaves (32). Genes involved in hormone pathways may serve as another example: The ZEP, AAO, and CYP707A families, all of which are involved in the abscisic acid (ABA) biosynthesis pathway, show an expansion of gene numbers in the *T. salsuginea* genome (*SI Appendix*, Table S9). This expansion may lead to a more complex regulation of ABA biogenesis, contributing to stress tolerance; the induction of gene expression by ABA in *Arabidopsis* is slower than in *T. salsuginea* until much higher stress levels have been reached (4, 6). The rapid ABA response in *T. salsuginea* under high salt conditions may confer a higher salinity-tolerance capacity by slowing down its growth rate. Further experiment evidence is necessary to confirm this hypothesis.

Additional salt stress-related gene families expanded in the *T. salsuginea* genome are summarized in *SI Appendix*, Table S10. Among them, *SAT32* is interesting because it is expanded from one gene in *A. thaliana* and *T. parvula* to six members in the *T. salsuginea* genome (*SI Appendix*, Fig. S6). *AtSAT32* is homologous to human IFN-related developmental regulator (IFRD) and is reported to be involved in salt-stress response (33). It is possible that these expanded family members give *T. salsuginea* more flexibility in response to salinity stress.

Discussion

With the increasing availability of second-generation sequencing, plant genome sequences are appearing in increasing numbers. Because of a desire to understand and improve agronomical important species, crops are an obvious target. A second group includes putative keystone species, i.e., models that might elucidate the evolutionary dimension of the genetic diversity essential to colonization of nearly every climate zone on earth. The third category is plants chosen for their close relationship to existing genomic and genetic models with the goal of expanding potential comparisons relevant at the biochemical and physiological level in particular environments. *T. salsuginea* is such a plant. On the one hand, it is phylogenetically and developmentally similar to the prototypical model, *A. thaliana*. It is a plant with halophytic characters and exceptionally high abiotic stress tolerance, including salinity, cold and freezing temperatures, and the ability to grow in poor soils. During the last decade it also has received considerable attention as a model of physiological and molecular defense against salinity stress (1–8). With the genome sequence presented in this study and with reference to the recently sequenced genome of the congeneric *T. parvula* (13), both juxtaposed with *Arabidopsis*, we have expanded the exploration of gene complement and allele structures that favored extremophile adaptations.

By tracing differences in genome structure and their evolutionary history, we have been able to point to processes that generated two species with extremely divergent adaptations within a time span of 7–12 million years (*SI Appendix*, Fig. S2). Although the gene spaces show extensive colinearity (Fig. 1), and the number of predicted gene models is similar to *A. thaliana* (*SI Appendix*, Fig. S1), selective expansions of seemingly stress-related gene families were observed in the *T. salsuginea* genome (Fig. 4 and *SI Appendix*, Tables S7–S10). Copy number variations of orthologs are largely caused by tandem and segmental duplication events that are unique to each species (Fig. 3*C*), similar to observations in the *T. parvula* genome (13).

However, the *T. salsuginea* genome is characterized by a dramatically higher content of TEs as compared with *A. thaliana* and *T. parvula*, and this greater number of TEs is largely responsible for its enlarged genome size (Fig. 1). Genes contained in LTR and retroelements are more abundant in *T. salsuginea*, with significantly higher numbers of these elements showing tandem duplications than in *A. thaliana* (455 vs. 307 genes). This observation confirms the role of TEs in tandem duplication events (Fig. 3 *A* and *B*). In addition, gene duplications have led to changes in gene dosage. Following sub- and neo-functionalization, functional diversification ensues, and duplicated copies that include favorable characters are retained in the process of natural selection. Duplicates lacking clear advantages for the organism turn into pseudogenes that eventually disappear (34). The stressful environment to which *T. salsuginea* has been exposed seems to have resulted in or to have contributed to the particular population of gene duplications that were retained. This view is supported by the presence of a comparable number but different suite of species-specific duplications in the *A. thaliana* genome (Fig. 3*C*). We also observed a number of translocation events for individual and small groups of genes relative to the *Arabidopsis* and *T. parvula* genomes, although it is not yet possible to assign a particular functional significance to these translocation events. Another outstanding character is the frequency of alterations in the sequences and *cis*-element structures of promoters for orthologous genes in the three species. These alterations can result in a complete rewiring of gene regulation, as exemplified by the expression of the duplicated *HKT1* genes (*SI Appendix*, Fig. S4) as well as for other stress-related genes (9).

Another level of complexity is present in the form of a substantial number of orphan genes that are specific to *T. salsuginea*. These genes do not have an ortholog in *A. thaliana* or in *T. parvula* (Fig. 2*B*) and frequently have no annotation based on sequence similarity. They may represent unique means of adaptation by providing domains with alternative functions, or they may be involved in shuffling of known protein domains. Compared with *Arabidopsis*, *T. salsuginea* is characterized by a dramatically different lifestyle, a unique gene complement, significant differences in the expression of orthologs, and a larger genome size. The *T. salsuginea* genome provides a tool for comparison and contrast to the well-established model, *Arabidopsis*. The resolution provided by the comparison between the two species is exceptionally high. Such resolution is not achievable by comparing plant genomes that are evolutionarily more distant. Multispecies comparative genomics strategies now can focus in detail on gene duplications in stress-related functions, neo-functionalization of duplicated genes, the consequences of translocation events, and orphan gene functions. The divergent regulation of gene expression in development and in communication with stressful environments now can be probed with the support of global transcript profiles. Because fundamental differences in handling stress are emerging (34), it seems that pathways and functions related to stress observed in *Arabidopsis* could be different in evolutionarily stress-tolerant plants. The genome of *T. salsuginea* will be a useful tool in exploring mechanisms of adaptive evolution.

Methods

DNA Library Construction and Sequencing. Short-insert DNA libraries (170 bp, 300 bp, and 800 bp) and long-insert DNA libraries (2 kb, 5 kb, and 10 kb) were built following protocols described previously (35). All libraries were subjected to paired-end sequencing runs, following the manufacturer's user guide (Illumina). A total of 12 DNA libraries were built and sequenced to ensure the randomness of clones. The raw sequence reads with base-calling duplicates, adapter contamination, PCR duplicates, and low-quality sequences were cleaned from the initial sequencing output using custom scripts.

Genome Assembly. We used a hybrid assembly and a hierarchical assembly approach with multiple assembly programs to build gap-free contigs, contigs combined to scaffolds, and scaffolds ordered into pseudochromosomes. At the first level of assembly, we used ABySS (36) and SOAPdenovo (35) followed by minimus2 (37) for meta-assembly of the primary contigs and scaffolds. Contigs were generated with a minimum of 10 overlapping high-quality mate pairs in stringent assembly parameters using 41- to 64-bp-long k-mers in search of high-quality contigs in the primary assemblies. Contigs with lengths less than 100 bp were discarded. In the initial assembly, 50 and 90% of the total length of 174,275,254 bp was covered by contigs larger than 3.23 kb and 149 bp, respectively (contig N50 = 3.23 kb, N90 = 149 bp). Scaffolds were assembled by adding all the paired-end reads to the initial contig assembly, followed by meta-assembly using minumus2. The assembled scaffolds were aligned to both the *T. parvula* (13) and *A. thaliana* (12) genome sequences using Nucmer (38). Scaffolds that could be aligned unambiguously to an ancestral karyotype block (39) in either *T. parvula* or *A. thaliana* were mapped to the karyotype model for the subclade *Eutremae* identified by comparative chromosome painting (10). Directions of mapped scaffolds were visualized further and corrected using the comparative genome visualization tool MAUVE (40), consulting both *T. parvula* and *A. thaliana* genomes. Scaffolds that could not be aligned unambiguously were labeled as unaligned.

More methods and details of data collection are provided in *SI Appendix*.

ACKNOWLEDGMENTS. We thank Beijing Genomics Institute staff members Ying Huang, Na An, Chunfang Peng, Yinqi Bai, Jianwen Li, Qingle Cai, Shiping Liu, Min Xie, Wei Fan, Bo Wang, Sheng Tang, Yuxiang Liu, Juan Wang, Kui Wu, Chuyu Lin, Yalin Huang, Kang Yi, Fei Teng, Fengjie Yu, Haibo Lin, Ruiqiang Li, Zhi Jiang, XiaoJu Qian, Hailong Luo, and Junjie Liu for their sequencing support. This research was supported by Grants 31030047, 30921061, 30825029, and 90917016 from the National Science Foundation of China, Grants 973 2012CB114300 and 2012CB114200 from the National Basic Research Program of China, and by Grant 2009A0714-05 from the State Key Laboratory of Plant Genomics of China. D.-H.O., S.Y.L., and H.J.B. are supported by World Class University Program R32–10148 at Gyeongsang National University, Republic of Korea and the Next-generation BioGreen21 Program SSAC, PJ009030, Rural Development Administration, Republic of Korea.

1. Amtmann A (2009) Learning from evolution: *Thellungiella* generates new knowledge on essential and critical components of abiotic stress tolerance in plants. *Mol Plant* 2:3–12.
2. Bressan RA, et al. (2001) Learning from the *Arabidopsis* experience. The next gene search paradigm. *Plant Physiol* 127:1354–1360.
3. Inan G, et al. (2004) Salt cress. A halophyte and cryophyte *Arabidopsis* relative model system and its applicability to molecular genetic analyses of growth and development of extremophiles. *Plant Physiol* 135:1718–1737.
4. Taji T, et al. (2004) Comparative genomics in salt tolerance between *Arabidopsis* and *aRabidopsis*-related halophyte salt cress using *Arabidopsis* microarray. *Plant Physiol* 135:1697–1709.
5. Wong CE, et al. (2006) Transcriptional profiling implicates novel interactions between abiotic stress and hormonal responses in *Thellungiella*, a close relative of *Arabidopsis*. *Plant Physiol* 140:1437–1450.
6. Gong Q, Li P, Ma S, Indu Rupassara S, Bohnert HJ (2005) Salinity stress adaptation competence in the extremophile *Thellungiella halophila* in comparison with its relative *Arabidopsis thaliana*. *Plant J* 44:826–839.
7. Taji T, et al. (2008) Large-scale collection and annotation of full-length enriched cDNAs from a model halophyte, *Thellungiella halophila*. *BMC Plant Biol* 8:115.
8. Wong CE, et al. (2005) Expressed sequence tags from the Yukon ecotype of *Thellungiella* reveal that gene expression in response to cold, drought and salinity shows little overlap. *Plant Mol Biol* 58:561–574.
9. Oh DH, et al. (2010) Genome structures and halophyte-specific gene expression of the extremophile *Thellungiella parvula* in comparison with *Thellungiella salsuginea* (*Thellungiella halophila*) and *Arabidopsis*. *Plant Physiol* 154:1040–1052.
10. Mandáková T, Lysak MA (2008) Chromosomal phylogeny and karyotype evolution in x=7 crucifer species (*Brassicaceae*). *Plant Cell* 20:2559–2570.
11. Lysak MA, et al. (2006) Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related *Brassicaceae* species. *Proc Natl Acad Sci USA* 103:5224–5229.
12. Initiative TAG; Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815.
13. Dassanayake M, et al. (2011) The genome of the extremophile crucifer *Thellungiella parvula*. *Nat Genet* 43:913–918.
14. Lysak MA, Koch MA (2011) Phylogeny, Genome, and Karyotype Evolution of Crucifers (*Brassicaceae*). *Genetics and Genomics of the Brassicaceae*, eds Schmidt R, Bancroft I (Springer, New York), pp 1–31.
15. Götz S, et al. (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* 36:3420–3435.
16. Fowler SG, Cook D, Thomashow MF (2005) Low temperature induction of *Arabidopsis* CBF1, 2, and 3 is gated by the circadian clock. *Plant Physiol* 137:961–968.
17. Sohn KH, Lee SC, Jung HW, Hong JK, Hwang BK (2006) Expression and functional roles of the pepper pathogen-induced transcription factor RAV1 in bacterial disease resistance, and drought and salt stress tolerance. *Plant Mol Biol* 61:897–915.
18. Lisso J, Altmann T, Müssig C (2006) The AtNFXL1 gene encodes a NF-X1 type zinc finger protein required for growth under salt stress. *FEBS Lett* 580:4851–4856.
19. Achard P, et al. (2008) The cold-inducible CBF1 factor-dependent signaling pathway modulates the accumulation of the growth-repressing DELLA proteins via its effect on gibberellin metabolism. *Plant Cell* 20:2117–2129.
20. Ogawa D, Yamaguchi K, Nishiuchi T (2007) High-level overexpression of the *Arabidopsis* HsfA2 gene confers not only increased themotolerance but also salt/osmotic stress tolerance and enhanced callus growth. *J Exp Bot* 58:3373–3383.
21. Fang Y, Xie K, Hou X, Hu H, Xiong L (2010) Systematic analysis of GT factor family of rice reveals a novel subfamily involved in stress responses. *Mol Genet Genomics* 283: 157–169.
22. Xie ZM, et al. (2009) Soybean Trihelix transcription factors GmGT-2A and GmGT-2B improve plant tolerance to abiotic stresses in transgenic *Arabidopsis*. *PLoS ONE* 4: e6898.
23. Volkov V, Amtmann A (2006) *Thellungiella halophila*, a salt-tolerant relative of *Arabidopsis thaliana*, has specific root ion-channel features supporting K^+/Na^+ homeostasis under salinity stress. *Plant J* 48:342–353.
24. Sun ZB, et al. (2008) Overexpression of a *Thellungiella halophila* CBL9 homolog, ThCBL9, confers salt and osmotic tolerances in transgenic *Arabidopsis thaliana*. *J Plant Biol* 51(1):25–34.
25. Lv S, et al. (2008) Overexpression of an H^+-PPase gene from *Thellungiella halophila* in cotton enhances salt tolerance and improves growth and photosynthetic performance. *Plant Cell Physiol* 49:1150–1164.
26. Vera-Estrella R, Barkla BJ, García-Ramírez L, Pantoja O (2005) Salt stress in *Thellungiella halophila* activates Na^+ transport mechanisms required for salinity tolerance. *Plant Physiol* 139:1507–1517.
27. Rus A, et al. (2006) Natural variants of AtHKT1 enhance Na^+ accumulation in two wild populations of *Arabidopsis*. *PLoS Genet* 2:e210.
28. Ren ZH, et al. (2005) A rice quantitative trait locus for salt tolerance encodes a sodium transporter. *Nat Genet* 37:1141–1146.
29. Byrt CS, et al. (2007) HKT1;5-like cation transporters linked to Na^+ exclusion loci in wheat, Nax2 and Kna1. *Plant Physiol* 143:1918–1928.
30. Ali Z, et al. (2012) TsHKT1;2, a HKT1 homolog from the extremophile *Arabidopsis* relative *Thellungiella salsuginea*, shows $K(^+)$ specificity in the presence of NaCl. *Plant Physiol* 158:1463–1474.
31. Kosma DK, et al. (2009) The impact of water deficiency on leaf cuticle lipids of *Arabidopsis*. *Plant Physiol* 151:1918–1929.
32. Teusink RS, Rahman M, Bressan RA, Jenks MA (2002) Cuticular waxes on *Arabidopsis thaliana* close relatives *Thellungiella halophila* and *Thellungiella parvula*. *Int J Plant Sci* 163(2):309–315.
33. Park MY, et al. (2009) Isolation and functional characterization of the *Arabidopsis* salt-tolerance 32 (AtSAT32) gene associated with salt tolerance and ABA signaling. *Physiol Plant* 135:426–435.
34. Oh DH, Dassanayake M, Bohnert HJ, Cheeseman JM (2012) Life at the extreme: Lessons from the genome. *Genome Biol* 13:241.
35. Li R, et al. (2010) The sequence and *de novo* assembly of the giant panda genome. *Nature* 463:311–317.
36. Simpson JT, et al. (2009) ABySS: A parallel assembler for short read sequence data. *Genome Res* 19:1117–1123.
37. Sommer DD, Delcher AL, Salzberg SL, Pop M (2007) Minimus: A fast, lightweight genome assembler. *BMC Bioinformatics* 8:64.
38. Kurtz S, et al. (2004) Versatile and open software for comparing large genomes. *Genome Biol* 5:R12.
39. Schranz ME, Lysak MA, Mitchell-Olds T (2006) The ABC's of comparative genomics in the Brassicaceae: Building blocks of crucifer genomes. *Trends Plant Sci* 11:535–542.
40. Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14:1394–1403.